Computers and the Humanities **37:** 97–109, 2003.

Analyzing the Order of Items in Manuscripts of *The Canterbury Tales*

MATTHEW SPENCER[1], BARBARA BORDALEJO[2], LI-SAN WANG[3], ADRIAN C. BARBROOK[1], LINNE R. MOONEY[4], PETER ROBINSON[2], TANDY WARNOW[3] and CHRISTOPHER J. HOWE[1]
[1]*Department of Biochemistry, University of Cambridge, UK*
E-mail: ms379@cam.ac.uk
[2]*Centre for Technology and the Arts, De Montfort University, UK*
E-mail: bb268@nyu.edu
[3]*Department of Computer Sciences, University of Texas, USA*
E-mail: lisan@cs.utexas.edu
[4]*Department of English, University of Maine, USA*
E-mail: mooney@maine.edu

Abstract. Chaucer's *Canterbury Tales* consists of loosely-connected stories, appearing in many different orders in extant manuscripts. Differences in order result from rearrangements by scribes during copying, and may reveal relationships among manuscripts. Identifying these relationships is analogous to determining evolutionary relationships among organisms from the order of genes on a genome. We use gene order analysis to construct a stemma for the *Canterbury Tales*. This stemma shows relationships predicted by earlier scholars, reveals new relationships, and shares features with a word variation stemma. Our results support the idea that there was no established order when the first manuscripts were written.

Key words: Canterbury Tales, Chaucer, gene order, phylogenetic analysis, stemmatology

1. Introduction

Chaucer's *Canterbury Tales* is a series of loosely-connected stories told by fictional pilgrims. Extant manuscripts of *The Canterbury Tales*, copied by hand in the fifteenth century, show many different orderings of the tales and linking passages (Manly and Rickert, 1940a). These differences are largely due to rearrangements of items (tales and links) by scribes, who found it difficult to establish an appropriate order even in the earliest manuscripts (Manly and Rickert, 1940a; Doyle and Parkes, 1979; Ramsey, 1994, pp. 307–343; Stubbs, 2000). The order of items may reveal genealogical relationships among manuscripts, but previous scholars have tried to discern these relationships without quantitative analysis (Manly and Rickert, 1940a; Dempster, 1949; Benson, 1981; Blake, 1985; Fisher, 1988; Owen, 1991; Ramsey, 1994). However, the transmission of item order from ancestral to

descendant manuscripts, with occasional modifications, is analogous to the transmission of the order of genes on a genome (Sankoff, 1992), and can be analyzed in a similar way. Here, we use methods developed for the analysis of gene order to produce a stemma based on the order of tales and links in *The Canterbury Tales*.

2. Methods

2.1. Item order data

We transcribed the order of tales in all reasonably complete extant manuscripts and early printed editions of *The Canterbury Tales* from Charts I–IV in Manly and Rickert (1940a). We recoded these data to include linking passages as well as tales. We treated homologous links (used, with slight alterations in their text, to connect different pairs of tales in different manuscripts) as the same items (treating them as distinct items gave similar results). We deleted one manuscript (Glasgow Hunterian 197) in which several items occur twice, and one manuscript (Northumberland 455) in which an item was split, because our distance measure does not deal with such cases. This left 56 manuscripts (Table 1) with 14 to 51 items each (median 44). The Hengwrt manuscript (Hg in Figure 1) was rebound in a slightly different order after production. We used the reconstructed original order of Hg (Manly and Rickert, 1940a) (using the rebound order gave similar results).

2.2. Distance measure

We estimated a matrix of pairwise evolutionary distances between item orders, where evolutionary distance is the expected number of transpositions (movements of one or more items) needed to turn one order into another. Insertions and deletions of items also occurred, but we cannot reconstruct these because leaves have been lost from many manuscripts. We therefore included only the items common to both manuscripts of a pair when estimating distances, and normalized all distances to the number of common items. We assumed that all transpositions were equally likely. This will not be strictly true, as there is strong internal evidence guiding the placement of some items such as the General Prologue, Knight's Tale and Miller's Tale. However, we are not able to estimate transposition probabilities from the data we have.

The most popular measure of evolutionary distance based on the order of items (such as genes on a genome) is breakpoint distance (Watterson *et al.*, 1982; Sankoff, 1992), the number of items whose right-hand neighbor is different between two orders. For example, the breakpoint distance between the hypothetical orders x and y

$$\begin{aligned} x &= 1\,2\,3\,4\,5 \\ y &= 1|3|2|4\,5 \end{aligned} \tag{1}$$

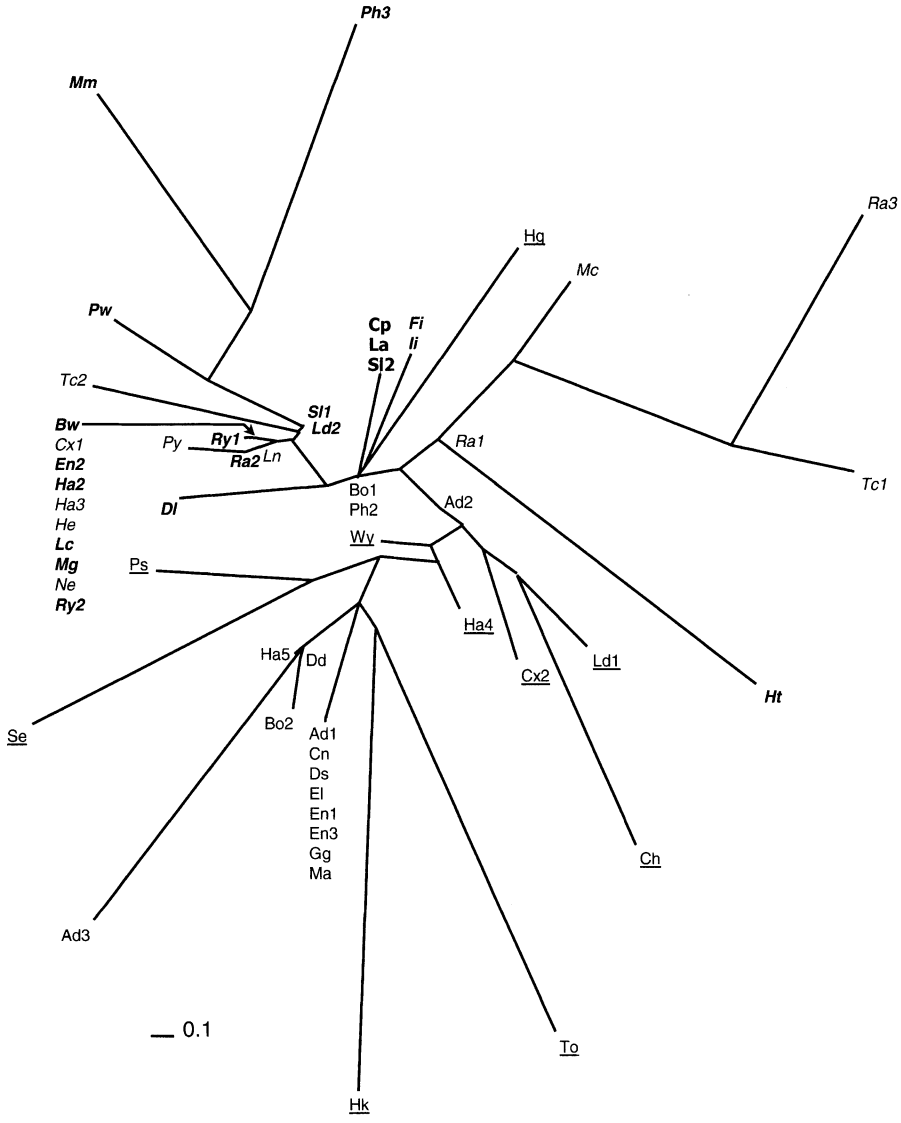

Figure 1. Unrooted minimum evolution stemma for reasonably complete extant manuscripts and early printed editions of *The Canterbury Tales* based on the order of tales and linking passages. Manuscripts are represented by the sigils given in Table 1, and the groups into which they were placed by Manly and Rickert (1940a) are indicated by fonts: "A" plain; "B" italic; "C" bold; "D" bold italic; "anomalous" underlined (Wy, Wynkyn de Worde's printed edition, is grouped as an anomalous witness, although it was not included in Manly and Rickert's data). Edge lengths are to scale, in units of IEBP distance. The sum of the lengths of edges connecting a pair of manuscripts is proportional to the number of transpositions needed to make their item orders identical. Manuscripts arranged in lists are located in visually indistinguishable positions on the stemma.

Table I. Information on the manuscripts and early printed editions included in this study. Names and locations are from the Sigils section of Solopova (2000). Approximate dates are from the Witnesses section of Solopova (2000), or from Manly and Rickert (1940b, volume II, pp. 46–48) for manuscripts lacking the General Prologue. Groups are from Manly and Rickert (1940a)

Sigil	Name/location	Date	Group
Manuscripts			
Ad1	London, British Library Additional MS 5140	s. XV^{ex}	A
Ad2	London, British Library Additional MS 25718	1430–1450	A
Ad3	London, British Library Additional MS 35286	s. XV$^{2/4}$	A
Bo1	Oxford, Bodleian Library MS Bodl. 414	s. XV^{2}	A
Bo2	Oxford, Bodleian Library MS Bodl. 686	s. XV$^{2/4?}$	A
Bw	Oxford, Bodleian Library MS Barlow 20	s. XV$^{3/4}$	D
Ch	Oxford, Christ Church MS 152	s. XV$^{3/4}$	anomalous
Cn	Austin, University of Texas HRC 43: Cardigan	1450?	A
Cp	Oxford, Corpus Christi College MS 198	s. XV^{in}	C
Dd	Cambridge University Library Dd.4.24	1401–1416	A
Dl	Tokyo, Takamiya MS 32: Delamere	s. XV$^{3/4}$	D
Ds	Tokyo, Takamiya MS 24: Devonshire	s. XV$^{3/4}$	A
El	San Marino, Huntington Library El. 26 C 9: Ellesmere	1400–1412?	A
En1	London, British Library Egerton MS 2726	s. XV$^{2/4}$	A
En2	London, British Library Egerton MS 2863	1430–1450	D
En3	London, British Library Egerton MS 2864	s. XV$^{4/4}$	A
Fi	Cambridge, Fitzwilliam Museum McClean 181	s. XV$^{3/4}$	D
Gg	Cambridge University Library Gg.4.27	s. XV$^{1/4}$	A
Ha2	London, British Library Harley MS 1758	s. XV$^{3/4}$	D
Ha3	London, British Library Harley MS 7333	Begun ca. 1450–1460	B
Ha4	London, British Library Harley MS 7334	s. XV^{in}	anomalous
Ha5	London, British Library Harley MS 7335	1450–1470	A
He	Princeton University Library MS 100: Helmingham	1420–1430 vellum, 50–60 paper	B
Hg	Aberystwyth, National Library of Wales Peniarth 392 D: Hengwrt	1400–1404	anomalous
Hk	Holkham Hall MS 667	1440–1450	anomalous
Ht	Oxford, Bodleian Library MS Hatton donat. 1	s. XV$^{3/4}$	D
Ii	Cambridge University Library Ii.3.26	s. XV$^{2/4}$	D
La	London, British Library Lansdowne MS 851	s. XV$^{1/4}$	C
Lc	Lichfield Cathedral MS 29	s. XV$^{2/4}$	D
Ld1	Oxford, Bodleian Library MS Laud Misc. 600	s. XV$^{3/4?}$	anomalous
Ld2	Oxford, Bodleian Library MS Laud Misc. 39	s. XV$^{4/4}$	D
Ln	Lincoln Cathedral Library MS 110	s. XV^{mid}	B

Table I. Continued

Sigil	Name/location	Date	Group
Ma	Manchester, John Rylands Library English MS 113	1483–1485	A
Mc	University of Chicago Library MS 564: McCormick	1440–1460	B
Mg	New York, Pierpont Morgan Library MS 249	s. XV^{mid}	D
Mm	Cambridge University Library Mm.2.5	s. XV^{mid}	D
Ne	Oxford, New College MS 314	s. $XV^{3/4}$	B
Ph2	Geneva, Bodmer Library 48	s. $XV^{3/4}$	A
Ph3	Philadelphia, Rosenbach Museum and Library 1084/1	1430–1450	D
Ps	Paris, Bibliothèque Nationale Fonds anglais 39	ca. 1422–1439	anomalous
Pw	Sussex, Petworth House MS 7	s. $XV^{2/4(early)}$	D
Py	London, Royal College of Physicians MS 388	s. $XV^{3/4}$	B
Ra1	Oxford, Bodleian Library MS Rawl. poet. 141	1450–1460	B
Ra2	Oxford, Bodleian Library MS Rawl. poet. 149	s. $XV^{3/4}$	D
Ra3	Oxford, Bodleian Library MS Rawl. poet. 223	s. $XV^{3/4}$	B
Ry1	London, British Library Royal MS 17 D.XV	s. $XV^{3/4}$	D
Ry2	London, British Library Royal MS 18 C.II	s. $XV^{2/4}$	D
Se	Oxford, Bodleian Library MS Arch. Selden. B. 14	s. $XV^{3/4}$	anomalous
Sl1	London, British Library Sloane MS 1685	s. $XV^{2/4}$	D
Sl2	London, British Library Sloane MS 1686	s. $XV^{4/4}$	C
Tc1	Cambridge, Trinity College R.3.3	s. $XV^{3/4}$	B
Tc2	Cambridge, Trinity College R.3.15	s. $XV^{4/4}$	B
To	Oxford, Trinity College MS 49	1461–1483	anomalous
Pre-1500 Printed Editions			
Cx1	Caxton, First Edition	c. 1477	B
Cx2	Caxton, Second Edition	c. 1483	anomalous
Wy	Wynkyn de Worde	1498	anomalous*

*Wy was not included in Manly and Rickert's groups, but is shown as "anomalous" on Figure 1.

is three (items are indicated by numbers and breakpoints by |). We could turn x into y by swapping the positions of items 2 and 3, so one transposition creates three breakpoints. However, the relationship between breakpoint distance and true evolutionary distance is nonlinear. We could further rearrange y into a new order z by swapping the positions of items 4 and 5, giving a sequence with a breakpoint distance of 4 from x:

$$\begin{aligned} x &= 1\,2\,3\,4\,5 \\ z &= 1|3|2|5|4 \end{aligned} \tag{2}$$

The first rearrangement (x to y) created three breakpoints, but the second rearrangement (y to z) added only one more. Thus breakpoint distance is only reliable when the number of transpositions is small.

To overcome this problem, we used the IEBP (Inverse of Expected BreakPoint distance) method (Wang and Warnow, 2001). IEBP (see the Appendix for details) estimates the most likely number of transpositions that occurred between a pair of orders, given the breakpoint distance between them. IEBP is better than breakpoint distance because it is an unbiased estimator of the actual number of transpositions. Given such an estimator, the methods we used to reconstruct stemmata (Section 2.3, below) are known to perform well (confirmed by simulation studies in Wang and Warnow, 2001). In any case, our results are robust because their main features remain the same even if breakpoint distance is used.

2.3. Stemma reconstruction

An item order stemma is a diagram showing how item order is transmitted from one manuscript to another. We assumed that such a stemma can be represented by a branching tree in which each manuscript has a single immediate ancestor. Nodes on the tree represent manuscripts, and the edges connecting these nodes represent the pathways of transmission of item order. The length of an edge is proportional to the IEBP distance between a pair of nodes, and the length of a tree is the sum of the lengths of its edges. For example, we often refer to the "shortest tree" as the tree with the smallest sum of edge lengths out of some set of trees. The distance on a tree between two manuscripts is the sum of the lengths of edges connecting the two manuscripts.

We reconstructed stemmata using the minimum evolution method implemented in the phylogenetic software package PAUP* (Swofford, 2001). Under the assumption that changes in item order are relatively rare, a short tree (requiring few changes of item order) is a more plausible explanation for an observed set of IEBP distances than a long tree. The minimum evolution method aims to find the shortest tree for a set of IEBP distances among manuscripts, where edge lengths for any given tree topology are estimated by minimizing the sum of squared deviations between the observed distances and distances on the tree. We constrained the algorithm to give edge lengths that were either zero or positive, because negative edge lengths do not have a meaningful interpretation. We set edges whose estimated lengths were less than 10^{-8} units of IEBP distance to zero, as such small distances were probably due to sampling error. Edges of length zero imply that a manuscript was copied more than once. The number of possible trees is very large (Flight, 1990; Li, 1997, pp. 102–103), so we did not evaluate them all. Instead, we started from an initially promising solution found by the neighbor-joining algorithm (Saitou and Nei, 1987) and attempted to improve it by removing branches from the tree and reconnecting them in different places.

2.4. Comparing item order and word variation stemmata

Variations in the texts of manuscripts can also reveal genealogical relationships (Robinson, 1997; Barbrook *et al.*, 1998). Congruence between stemmata based on different kinds of data increases our confidence that the stemmata are correct. The relationships among manuscripts based on word variation shift considerably between sections of text, because different sections of a manuscript may have been copied from different exemplars (Robinson, 1997) or may represent different stages of authorial revision (Moorman, 1993, pp. 74–79). However, a subset of 21 manuscripts has relatively consistent relationships over a short section of the text, the General Prologue (Solopova, 2000, Analysis Workshop). We extracted word variation data for the General Prologue in these 21 manuscripts from an electronic database (Solopova, 2000). We measured the distance between all pairs of manuscripts as mean character distance, the number of readings different between the manuscripts divided by the number of locations at which neither manuscript had missing text. This has a range from zero (the manuscripts are identical in all locations at which both are extant) to one (the manuscripts are different at all locations where both are extant). Mean character distance is a reasonable estimate of the number of changes of reading separating each pair of manuscripts, provided that it is unusual for several changes to occur at the same location in the text. This will be the case if all mean character distances are fairly small (Spencer and Howe, 2001), as they were in our data (the largest mean character distance between any pair of manuscripts was 0.23). We then reconstructed a minimum evolution stemma based on these distances using PAUP*, as described in section 2.3.

For the same subset of 21 manuscripts, the minimum evolution method found 30 item order stemmata with equally small sums of edge lengths. Since we have no reason to prefer any one of these stemmata over all the others, we summarized the information they had in common using an Adams-2 consensus stemma (Adams III, 1972; Swofford, 1991). Any pair of manuscripts that are grouped together in all of our 30 stemmata are grouped together on the consensus stemma. Areas where some of the stemmata disagree appear as unresolved sections in which many edges branch from the same node. We then applied the Adams-2 consensus method again to compare the stemma drawn from word variation and the consensus of the 30 equally good item order stemmata. This gives an overall graphical representation of similarity between the word variation and item order stemmata (the fewer unresolved sections, the higher the level of similarity).

We then used the partition metric (Penny and Hendy, 1985) to estimate the probability of observing random stemmata as similar as these word variation and item order stemmata. Any tree can be divided into two parts by removing any single edge. Each of the two parts of the tree contains a subset of the original set of manuscripts. We say that there is an equivalent edge in two trees containing the same manuscripts if we can find an edge a in each such that removing a creates exactly the same two subsets of manuscripts (taking no account of the

order in which manuscripts are arranged within each subset). The partition metric is simply the number of edges on one tree for which we can find no equivalent on the other tree. To obtain the distribution of the partition metric expected by chance, we generated 50000 random bifurcating trees using PAUP* and calculated the partition metric between each of these and the word variation stemma. For the observed distribution of the partition metric, we generated 50000 minimum evolution stemmata for the item order data, without setting the lengths of very short edges to zero, and calculated the partition metric between each and the word variation stemma. Not collapsing very short edges (less than 10^{-8} IEBP distance units) to zero results in a large number of trees differing only in trivial ways, and was necessary because it is difficult to generate suitable non-bifurcating trees at random.

We then compared the observed and random partition metric distributions using a Wilcoxon two-sample test corrected for ties (Sokal and Rohlf, 1995, p. 430). This is a non-parametric statistical test for differences in the location of two distributions. The data in both distributions (in this case, observed and random partition metrics) are combined and ranked, and the sum of ranks calculated for each. If one of the distributions has an unusually large sum of ranks, it indicates that most of the values from one distribution tend to be large relative to those from the other distribution.

3. Results

3.1. Item order stemma

A stemma based on IEBP distances (Figure 1) recovers many groups of manuscripts suggested by manual studies of tale order (Manly and Rickert, 1940a). Most manuscripts in the A (plain font) group are closely related, and the three manuscripts of the C (bold font) group have a common ancestor. The B (italic font) and D (bold italic font) groups are not clearly separated on our stemma. However, the order of items in all B and D manuscripts may have been derived from a single exemplar used for commercial copying (Dempster, 1949).

3.2. Comparing item order and word variation stemmata

Stemmata based on word variation (Figure 2A) and the order of tales and linking passages (Figure 2B) show some important similarities. The Adams-2 consensus between the word variation and item order stemmata (Figure 2C) retains groupings of manuscripts common to both stemmata, and shows disagreements as > 3 edges from a point. There are many fewer such disagreements than would be expected by chance (Wilcoxon two-sample test, $t_s = 128$, $P < 0.001$). For example, the A group (plain font) manuscripts are grouped together and the pairs {Tc2, Cx1} and {Ii, Ht} appear in both word variation (Figure 2A) and item order (Figure 2B) stemmata. However, the order of items and the text may have been transmitted

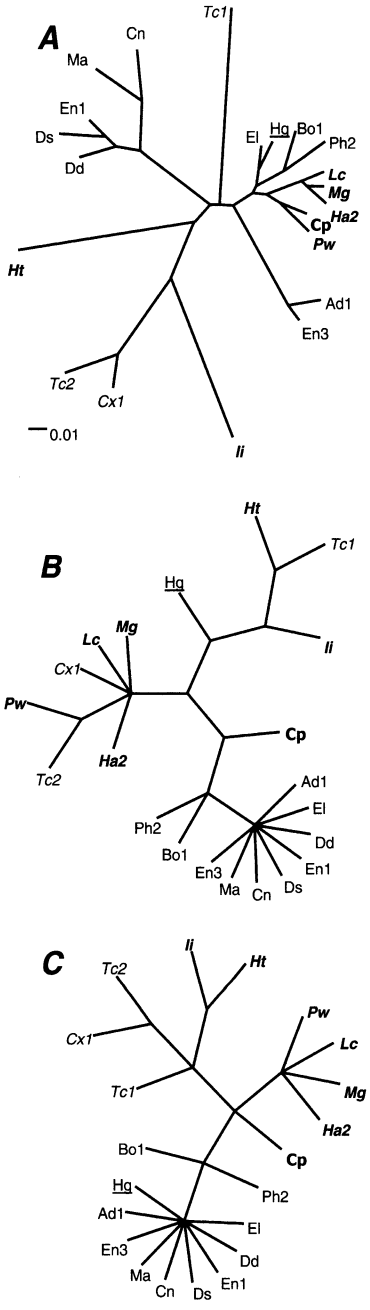

Figure 2. Comparisons between stemmata for *The Canterbury Tales* based on word variation and tale order. (A). Minimum evolution stemma based on word variation in lines 1–500 of the General Prologue, for 21 manuscripts with constant relationships. Edge lengths are to scale, in units of mean character distance. (B). Adams-2 consensus of 30 minimum evolution stemmata based on the order of tales and linking passages, in the same 21 manuscripts. Edge lengths not to scale. (C). Adams-2 consensus between the stemmata in (A) and (B). Edge lengths not to scale. Disagreements among candidate stemmata appear as > 3 edges from a point in the consensus. Sigils and fonts for groups as in Figure 1.

separately in some cases. For example, the manuscripts Hg and El were written by the same scribe (Doyle and Parkes, 1979) and have very similar texts in the General Prologue (Figure 2A), but have quite different orders (Figure 2B).

4. Discussion

The item order stemma (Figure 1) allows us to refine the conclusions of previous scholars. The manuscripts that were previously classified as "anomalous" (sigils underlined in Figure 1) (Manly and Rickert, 1940a) fall between the A (plain font) group and the other manuscripts. Cx2, Ch and Ld1 appear to have a common ancestor, as do To, Hk, Ps and Se. Analysis of the text of the anomalous manuscripts suggests that Cx2, Ha4 and Se may have been conflated from several different sources, while Ch and Hg may be close to the ancestor of the whole tradition, but independently descended from it (Robinson, 1997; Barbrook *et al.*, 1998). Hk appears to have been copied from a collection of fragments, and has different textual affiliation in different sections (Owen, 1991, p. 49; Robinson, 1997).

The stemma in Figure 1 is unrooted, so we cannot immediately determine which extant order is closest to the ancestor of the tradition. Even the earliest manuscripts (e.g. Hengwrt (Hg), Ellesmere (El), Cambridge Dd.4.24 (Dd), Corpus Christi Oxford 198 (Cp) and Harley 7334 (Ha4)), which we might expect to be close to the ancestor, are widely separated on the stemma. If Chaucer had a definite arrangement for the items, accurately represented in an extant manuscript, it would be hard to explain why the extant manuscripts have many different orders. Other literary works such as Boccaccio's *Decameron* and Gower's *Confessio Amantis*, produced around the same time as *The Canterbury Tales* (or a little earlier, in the case of the *Decameron*) and sharing a similar form, show little variation in the order of sections among extant manuscripts (Doyle and Parkes, 1978; Benson, 1981). This is consistent with Chaucer's original copy of *The Canterbury Tales* being unfinished and disordered (Blake, 1985, pp. 165–178; Fisher, 1988). Some sections may also have circulated independently during Chaucer's lifetime (Manly and Rickert, 1940a; Owen, 1991, p. 4; Ramsey, 1994, p. 285).

Elsewhere, we showed how small-scale patterns of word variation between manuscripts within a section of *The Canterbury Tales* can be analyzed using phylogenetic methods (Robinson, 1997; Barbrook *et al.*, 1998). Here, we establish a further parallel between the order of genes on a genome and the order of items in a text. Constructing a stemma based on the order of items has allowed us to hypothesize relationships among previously unclassified manuscripts, which can now be tested using data on word variation. Phylogenetic methods provide an objective visualization of the complex relationships among a set of genealogically-related items. These rigorous methods, combined with the increasing availability of manuscript data in electronic form, will change the way scholars approach literary traditions.

Acknowledgements

This work was supported by the Leverhulme Trust. We are grateful to Nick Colegrave, Jim Spencer, Rebecca Timms and Phil Warren for helpful suggestions, and to three anonymous reviewers for constructive criticism.

Appendix: Calculating IEBP distance

Here we outline the calculation of IEBP (Inverse of Expected BreakPoint distance) for a pair of linear orders rearranged only by transposition. See Wang and Warnow (2001) for a fuller exposition.

1. Compute the breakpoint distance b between two linear orders containing n items in common.

2. Let $s(i)$ be the probability of creating a new breakpoint at item i when a transposition occurs:

$$s(i) = \begin{cases} \frac{3(n-2)}{n(n-1)}, 0 < i < n \\ \frac{3}{n+1}, i = 0, i = n \end{cases} \tag{3}$$

Let $u_{\min}(i)$ and $u_{\max}(i)$ be the minimum and maximum probabilities of removing an existing breakpoint at item i when a transposition occurs, over all possible orders of items:

$$u_{\min}(i) = \begin{cases} \frac{6}{n(n-1)}, 0 < i < n \\ \frac{1}{\binom{n+1}{3}}, i = 0, i = n \end{cases} \tag{4}$$

$$u_{\max}(i) = \begin{cases} \frac{6}{n(n-1)}, 0 < i < n \\ \frac{6}{n(n+1)}, i = 0, i = n \end{cases} \tag{5}$$

3. We can then calculate lower and upper bounds $P^L_{i|k}$ and $P^H_{i|k}$ on the probability that there is a breakpoint at item i after k transpositions:

$$\begin{aligned} P^L_{i|k} &= s(i)\left(\frac{1-(1-s(i)-u_{\max}(i))^k}{1-(1-s(i)-u_{\max}(i))}\right) \\ P^H_{i|k} &= s(i)\left(\frac{1-(1-s(i)-u_{\min}(i))^k}{1-(1-s(i)-u_{\min}(i))}\right) \end{aligned} \tag{6}$$

4. Find the number of rearrangements k that minimizes the unsigned difference

$$\left|\sum_{i=0}^{n} \frac{P^L_{i|k} + P^H_{i|k}}{2} - b\right| \tag{7}$$

between the observed and expected number of breakpoints, summed over all n items. This is the IEBP distance. For the Canterbury Tales data, different pairs of manuscripts contained different numbers n of items in common, so we divided k by n to obtain a normalized IEBP distance.

References

Adams III E.N. (1972) Consensus Techniques and the Comparison of Taxonomic Trees. *Systematic Zoology*, 21, pp. 390–397.

Barbrook A.C., Howe C.J., Blake N., Robinson P. (1998) The Phylogeny of *The Canterbury Tales*. *Nature*, 394, p. 839.

Benson L.D. (1981) The Order of *The Canterbury Tales*. *Studies in the Age of Chaucer*, 3, pp. 77–120.

Blake N.F. (1985) *The Textual Tradition of The Canterbury Tales*. Edward Arnold, London, p. 222.

Dempster G. (1949) The Fifteenth-century Editors of the Canterbury Tales and the Problem of Tale Order. *Publications of the Modern Language Association of America*, 64, pp. 1123–1142.

Doyle A.I., Parkes M.B. (1978) The Production of Copies of the *Canterbury Tales* and the *Confessio Amantis* in the Early Fifteenth Century. In Parkes, M.B. and Watson, A.G. (eds.), *Medieval Scribes, Manuscripts and Libraries, Essays Presented to N.R. Ker*, Scolar, London, pp. 163–210.

Doyle A.I., Parkes M.B. (1979) Paleographical Introduction. In Ruggiers, P.G. (ed.), *The Canterbury Tales. Geoffrey Chaucer. A Facsimile and Transcription of the Hengwrt Manuscript*, University of Oklahoma Press, Norman, pp. xix–xlix.

Fisher J.H. (1988) Animadversions on the Text of Chaucer. *Speculum*, 63, pp. 779–793.

Flight C. (1990) How many stemmata? *Manuscripta*, 34, pp. 122–128.

Li W.-H. (1997) *Molecular Evolution*. Sinauer Associates, Sunderland, MA, p. 487.

Manly J.M., Rickert E. (1940a) The Order of Tales. In Manly, J.M. and Rickert, E. (eds.), *The Text of The Canterbury Tales: Studied on the Basis of All Known Manuscripts*, University of Chicago Press, Chicago, pp. 475–494.

Manly J.M., Rickert E. (eds.) (1940b) *The Text of the Canterbury Tales: Studied on the Basis of All Known Manuscripts*. University of Chicago Press, Chicago.

Moorman C. (1993) *The Statistical Determination of Affiliation in the Landmark Manuscripts of The Canterbury Tales*. The Edwin Mellen Press, Lewiston, p. 209.

Owen C.A., Jr. (1991) *The Manuscripts of The Canterbury Tales*. D.S. Brewer, Cambridge, p. 132.

Penny D., Hendy M.D. (1985) The Use of Tree Comparison Metrics. *Systematic Zoology*, 34, pp. 75–82.

Ramsey R.V. (1994) *The Manley-Rickert Text of the Canterbury Tales*. The Edwin Mellen Press, Lewiston, NY, p. 691.

Robinson P. (1997) A Stemmatic Analysis of the Fifteenth-century Witnesses to The Wife of Bath's Prologue. In Blake, N. and Robinson, P. (eds.), *The Canterbury Tales Project: Occasional Papers Volume II*, Office for Humanities Communication Publications, London, pp. 69–132.

Saitou N., Nei M. (1987) The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution*, 4, pp. 406–425.

Sankoff D. (1992) Edit Distance for Genome Comparison Based on Non-local Operations. *Lecture Notes in Computer Science*, 644, pp. 121–135.

Sokal R.R., Rohlf F.J. (1995) *Biometry*. W.H. Freeman & Co., New York, p. 887.

Solopova E. (ed.) (2000) *Geoffrey Chaucer: The General Prologue on CD-ROM*. Cambridge University Press, Cambridge.

Spencer M., Howe C.J. (2001) Estimating Distances between Manuscripts Based on Copying Errors. *Literary and Linguistic Computing*, 16, pp. 467–484.

Stubbs E. (2000) Observations on the Hengwrt Chaucer. In Stubbs, E. (ed.), *The Hengwrt Chaucer Digital Facsimile*, Scholarly Digital Editions, Leicester.

Swofford D.L. (1991) When are Phylogeny Estimates from Molecular and Morphological Data Incongruent? In Miyamoto, M.M. and Cracraft, J. (eds.), *Phylogenetic Analysis of DNA Sequences*, Oxford University Press, Oxford, pp. 295–333.

Swofford D.L. (2001) *PAUP**. *Phylogenetic Analysis Using Parsimony* (*and other methods). Sinauer Associates, Sunderland, MA.

Wang L.-S., Warnow T. (2001) *Estimating True Evolutionary Distance between Genomes. Proceedings of the Thirty-Third Annual ACM Symposium on the Theory of Computing* (STOC'01), Crete, Greece, ACM Press, pp. 637–646.

Watterson G.A., Ewens W.J., Hall T.E. (1982) The Chromosome Inversion Problem. *Journal of Theoretical Biology*, 99, pp. 1–7.